©Journal of Aquatropica Asia
e-issn: 2721-7574; p-issn: 2407-3601
Volume 10, Nomor 2, Tahun 2025
Jurusan Akuakultur, Universitas Bangka Belitung

OPTIMIZATION OF A FLAT-DIE PELLETIZING MACHINE FOR FISH FEED PRODUCTION USING MACHINE LEARNING MODELS

OPTIMASI MESIN PENCETAK PELET TIPE FLAT-DIE UNTUK PRODUKSI PAKAN IKAN MENGGUNAKAN MODEL PEMBELAJARAN MESIN

Adedipe J.O, Yusuf I.O, Ibiyeye D.E

Federal College of Forestry, P.M.B. 5087 Jericho, Ibadan, Oyo State, Nigeria

**Corresponding author's email: crowndipe04@gmail.com*

Abstract

High-quality fish feed production is a critical factor in the success of aquaculture systems. The quality of feed pellets, particularly their durability (pellet durability index), directly affects feed efficiency and aquatic environmental quality. This study applied machine learning approaches to predict and optimize pellet durability index in a flat die pelletizing machine used for fish feed production. The dataset consisted of 3,000 observations with 13 operational features collected through IoT sensors. Descriptive statistics were used to summarize overall data characteristics. The target variable was binarized to classify pellets as acceptable or non-acceptable. To address data imbalance, the Synthetic Minority Oversampling Technique (SMOTE) was applied, followed by data cleaning and feature scaling to ensure uniformity. The preprocessed dataset was split into training and testing subsets. Six machine learning models were developed and evaluated using four performance metrics. Logistic Regression and Support Vector Machine achieved the best performance, with accuracy, recall, precision, and F1-score values of 0.756, 1.000, 0.756, and 0.861, respectively. The results indicate that operational factors in the fish feed pelletization process can effectively train predictive models to assess feed quality. The integration of machine learning in fish feed manufacturing offers potential improvements in production efficiency and feed quality for commercial aquaculture applications.

Keywords: Fish feed, Pellet durability, Machine learning, Smart aquaculture, IoT, Feed quality prediction.

Abstrak

Produksi pakan ikan yang berkualitas tinggi merupakan faktor penting dalam keberhasilan budidaya perikanan. Kualitas pellet pakan, khususnya daya tahannya (pellet durability index), sangat berpengaruh terhadap efisiensi pakan dan kualitas lingkungan perairan. Penelitian ini menerapkan pendekatan machine learning untuk memprediksi dan mengoptimalkan indeks daya tahan pellet pada mesin flat die pelletizer yang digunakan dalam produksi pakan ikan. Dataset yang digunakan terdiri dari 3.000 observasi dengan 13 fitur operasional yang diperoleh melalui sensor IoT. Analisis deskriptif digunakan untuk memperoleh gambaran umum data. Variabel target dibinarisasi untuk mengklasifikasikan pellet menjadi kategori "diterima" dan "tidak diterima". Ketidakseimbangan data diatasi dengan Synthetic Minority Oversampling Technique (SMOTE), sementara data cleaning dan feature scaling dilakukan untuk meningkatkan kualitas data. Dataset kemudian dibagi menjadi data latih dan uji. Enam model machine learning dievaluasi menggunakan empat metrik performa. Hasil menunjukkan bahwa Logistic Regression dan Support Vector Machine memiliki performa terbaik dengan nilai akurasi, recall, presisi, dan F1-score masing-masing 0,756; 1,000; 0,756; dan 0,861. Temuan ini menunjukkan bahwa faktor-faktor proses produksi pellet dapat digunakan untuk melatih model prediktif yang andal dalam menentukan kualitas pakan ikan. Penerapan machine learning pada proses manufaktur pakan ikan berpotensi meningkatkan efisiensi produksi dan mutu pakan dalam industri akuakultur komersial.

Kata Kunci: Pakan ikan, Daya tahan pellet, Pembelajaran mesin, Akuakultur cerdas, IoT, Prediksi kualitas pakan.

INTRODUCTION

Fish farming, or aquaculture, relies heavily on the quality of feed to ensure healthy and efficient fish growth (Hasna et al., 2018). Pellets used in fish feed must meet specific requirements such as water stability, durability, and nutrient retention. Flat die pellet mills are widely used in small and medium-scale fish feed production due to their compact design and ease of use. Pellet quality can be defined as the capacity of pelleted feeds to withstand fragmentation and abrasion during handling and transport, and maintain their integrity (Muramatsu et al., 2015). The quality of pellets is measured by durability, density, and uniformity. It also depends on operational parameters like motor power, feed rate, and moisture content. Traditionally, optimizing these parameters relies on manual adjustments or heuristic methods, which can be inefficient and imprecise (Khater et al., 2014).

Recent studies have underscored the importance of optimizing pelletizing parameters. However, optimizing the pelletizing process to consistently produce high-quality feed is challenging. (Smith et al., 2015; Jones and Brown, 2021; Buchanan and Moritz, 2009; Muramatsu et al., 2015) highlight that variables such as compression ratio, ingredients particle size and moisture content significantly influence pellet density, bonding and durability.

Machine learning (ML) provides a robust solution by enabling predictive analytics for product quality. They are suitable for prediction tasks (Bzdok et al., 2018). The application of ML in manufacturing processes has gained traction. Zhang et al. (2023) used ML to optimize injection molding, achieving a 12% improvement in product quality by predicting optimal parameter settings. In pelletizing, ML has been underutilized, though preliminary studies, Kim (2024) suggests that ML models like Random Forests and Neural Networks can predict pellet durability with over 90% accuracy. These models excel at handling non-linear relationships between parameters, such as the interplay between moisture content and pressure. The aim of the research is to compare different ML models using real time parameters for the optimization of a flat die pelletizer.

Research Gap

Recently, ML models such as Artificial Neural Networks (ANN), Decision Trees (DT), and Support Vector Machines (SVM) have been employed to predict pellet quality. Despite these advancements, few studies have applied ML to pelletizing for real-time optimization.

Objective

This research addresses this gap by; Developing and comparing different ML models using real time parameters and predict Pellet Durability Index and identify optimal operating parameters.

MATERIALS AND METHODS

Materials

The following materials were used in this study; flat die pelletizer, measuring device for real time features, and pelletizing ingredients. The ingredients were used to prepare the formulation for the fish pellets; protein sources (fish and soybean meal), carbohydrate (maize and wheat bran), fats and oil (fish oil), binders (cassava starch) and vitamin (dicalcium phosphate).

Machine Description

The pelletizing machine used in this study was a flat die pellet mill, designed for processing fish feed ingredients into pellets. It consists of a flat circular die that is stationary and above it is rollers that press the materials into the die holes. It has a feeding mechanism that moves the material and a cutter that trims the pellets into uniform sizes. It also consists of rollers, feed system, and drive motor. The design ensures appropriate compression, which is essential for water-stable and durable pellets required in aquaculture.

Methods

The selected ingredients were weighed using a precision scale according to a pre-determined feed formulation and grinded to reduce particle sizes. The dry and wet ingredients were thoroughly mixed for homogeneity and then steamed to soften and improve binding. A moisture meter was used to measure the moisture content. The conditioned mash was then passed through a pellet mill die to produce uniform pellets. Figure 1 shows a workflow diagram for the research. All data analysis work was performed using Python and supporting libraries like NumPy, Scikit-learn, Matplotlib, and SciPy (Harris et al., 2020). The Matplotlib library was used for data visualization, and the others were used for data cleaning, data pre-processing, modeling, and model evaluation.

Data Collection

A flat-die pellet mill was used to compress raw material into pellet while real- time operational data was collected from multiple production runs under varying condition using sensor data measuring devices and categorized into input and output parameters as shown in table 1. There was a total of 3000 observations corresponding to respective batch of pelleted feed. It comprised of 14 features in the dataset comprising of 10 continuous inputs, 3 continuous outputs and 1 categorical output.

Data Preprocessing

For this study, there were 13 features and 1 target feature (Pellet Durability Index). Pellet Durability Index (PDI) was determined as a percentage of the weight of the pellets measured after tumbling to the weight of the pellets before tumbling. The describe method of pandas DataFrame was used to estimate the overall statistics about the data as presented in table 2. The binarization of the target variable was carried out to classify the values into "1" acceptable and "0" non-acceptable. The data was then tested for imbalance using the synthetic minority oversampling technique (SMOTE) to ensure balance. Data cleaning was carried out to remove outliers and handle missing values. Feature scaling was carried out to put all features into the same scale using StandardScaler. The clean pre-processed dataset (3000 rows and 14 columns) was split uniformly at random into two subsets; a training set containing 80 % and a testing set 20% of the data. The training data set was used for training and validation of the models, while the testing set was used for the final evaluation of the models.

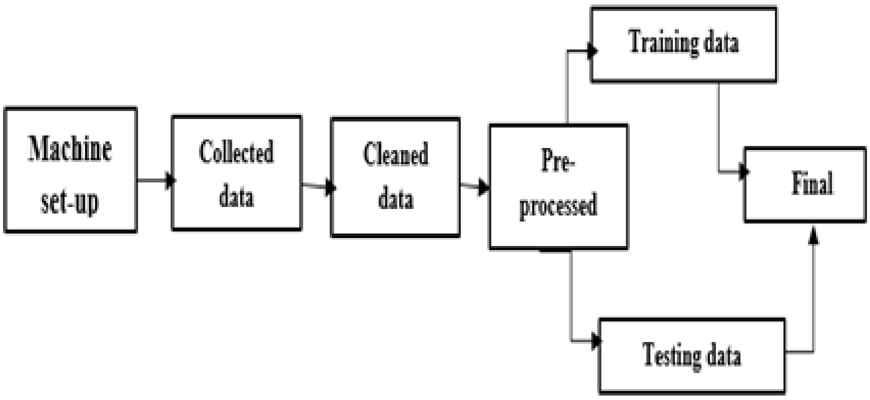

Figure 1. Workflow for the research

Table 1. Determined features and their measuring devices

No	Feature	Measuring device	Category
1	Energy	Energy Meter	Output
2	Throughput	Weight Sensor	Output
3	Moisture Content	Moisture Meter	Input
4	Die Size Hole	Digital Calipers	Input
5	Bulk Density	Volume over density	Input
6	Pressure	Pressure Sensor	Input
7	Temperature	Thermometer	Input
8	Motor Speed	Tachometer	Input
9	Feed Rate	Mass Flow Sensors	Input
10	Compression Ratio	Optical Sensors	Input
11	Pellet Durability Index	Tumbler Drum	Output
12	Wear Rate	Vibration Sensors	Output

RESULTS

Model Features

The influence of features on PDI prediction were estimated using the permutation importance approach (Breiman, 2001) since only a handful of ML algorithms provide an inner mechanism to estimate it (e.g., RF and DT). It has the ability to correct tree-based models since it can correct the bias from the impurity-based feature importance (Altmann et al., 2010). Therefore, the permutation importance was performed iteratively on every feature of the model and the prediction results were compared against the original predictions.

Model Development

The Python scikit-learn library (Pedregosa et al., 2011) was used to implement 6 ML different classifier algorithms which include; Decision Tree, Gradient Boosting Regression, Random Forest, Logistic Regression, K-Nearest Neighbor, Support Vector Classifier. These classification models were used to classify pellet durability index into "acceptable" or "non-acceptable". All algorithms, the corresponding acronyms, and the corresponding Python functions from the scikit-learn library are described in table 3.

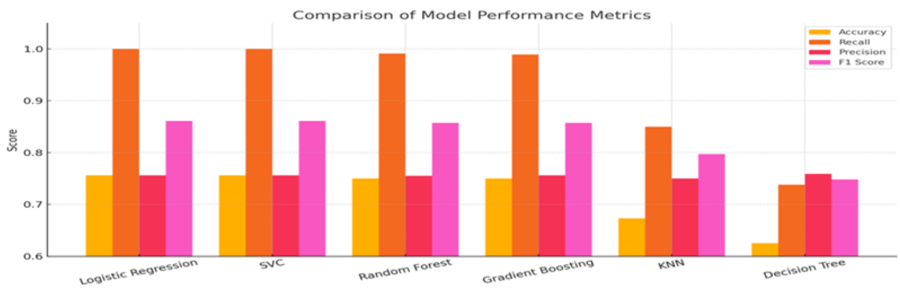

Figure 2. Comparison of model performance metrics

Table 2. Statistical description features

Feature	Count	Mean	Std	Min	25%	50%	75%	Max
Moisture content (%)	3000	12.99	2.89	8	10.52	12.98	15.48	17.99
Particle size (mm)	3000	2.77	1.31	0.5	1.63	2.78	3.9	5
Bulk density (kg/m^3)	3000	500.16	116.47	300.09	398.38	408.9	600.54	699.68
Temperature ^{0}C	3000	94.92	14.54	70.03	82.08	94.84	107.66	119.99
Pressure (Mpa)	3000	1.75	0.74	0.5	1.09	1.79	2.4	3
Die size (mm)	3000	5.97	2.3	2	3.97	6	7.96	10
Motor speed (rpm)	3000	1758.18	730.84	501.2	1114.38	1766.56	2405.28	2998.38
Feed rate (kg/h)	3000	276.46	129.08	50.24	164.3	279.13	388.03	499.57
Compression ratio	3000	10.03	2.87	5.01	7.51	10.11	12.5	15
Pellet density (kg/m^3)	3000	950.44	203.09	600.33	774.16	946.95	1123.79	1299.7
Pellet durability index	3000	89.49	5.51	80	84.72	89.40	94.24	98.98
Energy consumption (KWH)	3000	27.5	12.99	5.01	16.28	27.43	38.58	49.96
Throughput (kg/h)	3000	239.63	120.08	30.02	136.98	239.92	344.78	499.96
Wear rate (%)	3000	2.57	1.41	0.1	1.37	2.55	3.78	5

Table 3. The classifier algorithms used for model development

	Model	Acronym Function in the scikit-learn library
1	Logistic Regression (LR)	sklearn.linear_model.LogisticRegression
2	Support Vector Classifier (SVC)	sklearn.svm.SVC
3	K neighbors Classifier (KNC)	sklearn.neighbors.KNeighborsClassifier
4	Decision Tree Classifier (DTC)	sklearn.tree. DecisionTreeClassifier
5	Random Forest Classifier (RFC)	sklearn.ensemble.RandomForestClassifier
6	Gradient Boosting Classifier (GBR)	sklearn.ensemble

Table 4. Ranking of models by performance

Position	Model
1st	Logistic Regression
2nd	Support Vector Classifier
3rd	Random Forest Classifier
4th	Gradient Boosting Classifier
5th	K-Nearest Neighbor
6th	Decision Tree Classifier

Model Evaluation

Four statistical methods (Accuracy, precision, recall, and F1-score) were employed in this study to measure the performance of the classification models and the results were ranked based on their performance. The equations for respective methods are provided below.

Accuracy score = (TP+TN)/(TP+TN+FP+FN) (1)

Recall score = TP / (TP + FN) (2)

Precision score = TP / (TP + FP) (3)

$$\text{F1 score} = \frac{2\ (\text{Precision x Recall})}{(\text{Precision + Recall})} \quad (4)$$

Where TP = true positive, TN = true negative, FP = false positive, FN = false negative.

DISCUSSION

Model Performance

Figure 2 compares the performance of the six machine learning models based on four evaluation metrics; Accuracy, Recall, Precision, and F1 Score. Logistic Regression (LR) correctly predicted 75.6% of all cases, a perfect recall (1.000) indicates that the model successfully identified all the "acceptable" PDI values. However, the precision of 0.756 suggests that not all predicted "acceptable" values were actually acceptable, implying some false positives. The F1-score, which balances recall and precision, is relatively high (0.861), indicating a robust trade-off.

Support Vector Classifier (SVC) shared identical performance metrics with LR. This suggests that both models classified the test data similarly, likely due to a linearly separable decision boundary in the feature space. Random Forest Classifier (RFC) had an accuracy of 75%, a nearly perfect recall of 0.991 and maintained a good precision of 0.755 that led to a solid F1-score of 0.857. This means the model rarely misses an "acceptable" pellet, though a few falser positive predictions than LR/SVC. Gradient Boosting Classifier (GBC) closely mirrors RFC. It was slightly behind in recall (0.989 against 0.991) but achieved the same F1-score. This means it is also effective in identifying "acceptable" pellets but with slight variation in predictions.

K-Nearest Neighbors (KNC) model underperformed relative to the top four, with a lower accuracy of 0.673, F1-score of 0.79 and had a drop in recall of 0.850 and a precision of 0.750. Decision Tree Classifier (DTC) had the lowest overall performance with accuracy of 0.625. Despite decent precision value 0f of 0.759, it had a low recall of 0.738. Its poor F1-score of 0.748 also reflects a weak balance between precision and recall.

LR and SVC showed the highest performance in terms of recall and F1 score, while RFC and GBC also performed well. KNC and DT showed lower scores across most metrics, indicating less reliability for predicting pellet durability.

Overall Ranking Of Models By Performance

Based on F1-score, which is the most balanced indicator when both false positives and false negatives are important, table 3 further illustrates the performance of the models by their ranks. Although LR and SVC had the same F1-score, the final choice depended on ease of interpretation which favors LR. Given the performance metrics and practical implications, the Logistic Regression model was the most appropriate for the study due to its perfect recall score (no missed "acceptable" pellets), competitive precision and F1-score, it is simple, transparent, and can easily be integrated with real-time feedback systems for optimization.

Implications For Aquaculture

Pellet durability ranged from 75% to 95%, with higher values associated with lower moisture content and higher compression ratios. This finding agrees with Smith et al. (2015) that low moisture content improves pellet binding and Jones and Brown (2021) who highlighted the impact of compression ratio on pellet density. Optimized feed pellets would improve feeding efficiency, reduce waste, and enhance fish health.

CONCLUSION

This study demonstrates the effectiveness of machine learning in optimizing the operation of a flat die pelletizing machine for fish feed production by accurately predicting the pellet durability index (PDI). Among the six models tested, Logistic Regression and Support Vector Classifier achieved the best performance, both recording an accuracy of 75.6%, a perfect recall of 1.000, and an F1-score of 0.861. These results indicate that machine learning algorithms can reliably predict pellet quality based on operational parameters. Logistic Regression is particularly suitable for real-time deployment due to its simplicity, interpretability, and computational efficiency. The study concludes that factors related to fish feed manufacturing can be effectively utilized to develop predictive models for assessing pellet quality in commercial feed mills. Integrating such models within feed production systems provides valuable decision-support tools for optimizing feed formulation and processing. This approach supports the development of smart aquaculture by improving feed production efficiency, ensuring consistent pellet quality, and promoting sustainable fish farming practices.

REFERENCES

Altmann, A., Tolos¸i, L., Sander, O., Lengauer, T. (2010). Permutation importance: a corrected feature importance measure. Bioinformatics 26, 1340–1347. https://doi.org/10.1093/bioinformatics/btq134.

Breiman, L., (2001). Random forests. Mach. Learn 45, 5–32. https://doi.org/10.1023/A:1010933404324.

Buchanan, N.P., Moritz, J.S. (2009). Main effects and interactions of varying formulation protein, fiber, and moisture on feed manufacture and pellet quality. J. Appl. Poult. Res. 18, 274–283. https://doi.org/10.3382/japr.2008-00089.

Bzdok, D., Altman, N., Krzywinski, M. (2018). Statistics versus machine learning. Nat. Methods 15, 233–234. https://doi.org/10.1038 /nmeth.4642.

El-Sayed G Khater, Adel H Bahnasawy, Samir A Ali (2014). Physical and Mechanical Properties of Fish Feed Pellets. J Food Process Technol. 5:378.

Harris, C.R., Millman, K.J., van der Walt, S.J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N.J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M.H., Brett, M., Haldane, A., del Río, J.F., Wiebe, M., Peterson, P., G´erard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., Oliphant, T.E. (2020). Array programming with NumPy. Nature 585, 357–362. https://doi.org/10.1038/s41586-020-2649-2.

Hasna Jahan, Imran Parvez, Abu Syed Md. Kibria (2018). Current aquaculture practices in Dinajpur District Special emphasis on fish feeds. International Journal of Fisheries and Aquatic Research, 3(1), 35-42.

Jones, A. and Brown, K. (2021). Compression Ratio and Pellet Quality. Industrial Engineering Journal, 52(4), 200–210.

Kim, H. (2024). Predicting Pellet Durability with ML. Pellet Technology Journal, 18(2), 67–73.

Muramatsu, K., Massuquetto, A., Dahlke, F., Maiorka, A.2015). Factors that affect pellet quality: a review. J. Agric. Sci. Technol. 9, 717–722. https://doi.org/ 10.17265/2161-6256/2015.09.002. Noble, W.S., 2006. What is a support vector machine? Nat. Biotechnol. 24, 1565–1567. https://doi.org/10.1038/nbt1206-1565.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., (2011). Scikit-learn: machine learning in Python. J. Mach. Learn. Res. 12, 6.

Smith, J., Johnson, A., and Lee, K (2015). Moisture Effects on Biomass Pelletizing. Journal of Energy Research, 12(3), 123–130.

Lee, S. (2022). Vibration Monitoring in Pellet Mills. Mechanical Systems and Signal Processing, 38(2), 90–98.

Zhang C., Wu, Y., Zhang, W., and Yu, Y (2023). ML in Manufacturing Optimization. IET Collaborative intelligent Manufacturing, 10 (1), 45–52.